Toward a Common Earth Data Publication Framework

EOSDIS NASA'S EARTH OBSERVING SYSTEM DATA AND INFORMATION SYSTEM

AGU:IN53E-0651

Justin Rice[1], Daine Wright[2], Deborah Smith[3], Amanda Leon[4], Ajinkya Kulkarni[3]

Abstract

Data publication is an essential activity for all data archives. Each of NASA's twelve Distributed Active Archive Centers (DAACs) have established publication workflows which account for the heterogeneous suite of missions, instruments, data providers, and datasets managed within the Earth Observation System Data and Information System (EOSDIS) program. Some aspects of data publication vary across DAACs: workflows range from manual to automatic, terms used to describe publication elements differ, and systems used to publish and manage data vary. Despite these differences, the DAAC data publication processes are generally the same: obtain the data and related information from data providers, describe the data with metadata and documentation, release the data for access by the user community, archive and perform routine tests on the data, and provide support for the data and user community.

In order to improve consistency and reduce the time required to publish data, we have developed a cross-DAAC initiative called the Common Earthdata Publication Framework (Earthdata Pub). Earthdata Pub seeks to: standardize communications and interactions with data providers; identify and standardize common workflows and steps in the data publication process; and design/implement a front-end system with features that include a common web interface, email & status tracking, and common application programming interfaces (APIs) to communicate with various DAAC-specific software components (services and applications) on the back-end. This poster presents the latest updates on this effort's progress and future plans.

Motivations

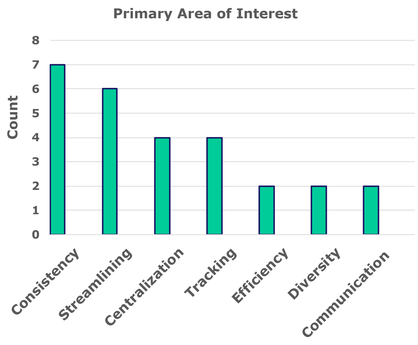
All DAACs follow data publication workflows to ingest and describe data for public distribution.

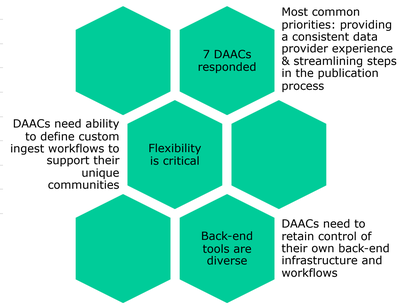
Receive data & info from data providers → Verify data → Verify/create metadata → Curate documentation → Release to public

Workflows vary from DAAC to DAAC. Some differences are inherent to the different data sources, datasets, and data users that the DAACs support or the data systems they operate.

There are opportunities to improve consistency and efficiency.

DAAC Survey Results





Earthdata Pub Goals

- **Assess and document** — to understand how DAACs communicate with data providers
- **Provide resources** — that improves communication channels
- **Develop software framework** — that creates a consistent experience for data providers across DAACs and reduces time required to publish data

Division of Labor

Information Group
- Identify information requirements for data publication process

Development Group
- Identify software framework requirements. Design and implement software framework.

Group Synergy
- Informational requirements identified by the information group will help inform some of the functional and technical requirements for the development group

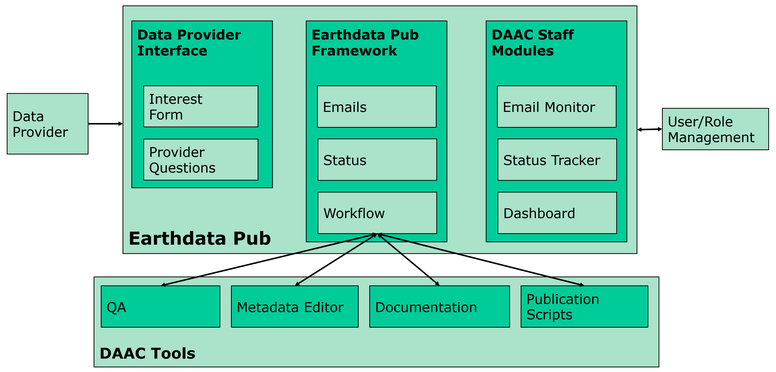
Group Goals

Information Group	Development Group
Identify commonalities among existing information resources	Identify and select common publication elements and tools
Standardize communications and interactions with data providers	Survey existing workflow management tools
Consistently communicate publication process	Create software project plan
Provide introductory material to first-timers	Develop software framework and backbone
Identify common information needed by DAACs from data providers and vice versa	Develop modules for information exchange forms

Progress

Information Group	Development Group
Developed and disseminated data publication survey to DAACs	Surveyed existing workflow management tools
Created a list of data publication pain points	Developed use cases on how DAACs will use software framework
Collected existing data publication resources	Finalized project plan that provides technical detailed information on the software framework

Software Framework Phase I



Future Work

Information Group	Development Group
Report on common DAAC data publication practices	Determine how to divide the labor for developing shared modules
Complete the terminology synonym list	Shore up details regarding API development
Complete the data provider "getting started guide"	Shore up details regarding retrofitting existing tools
Provide development group with information needed for software framework	Begin development of software framework and modules

Authors

[1] **NASA Goddard Space Flight Center**
[2] **Oak Ridge National Laboratory**
[3] **University of Alabama in Huntsville**
[4] **National Snow and Ice Data Center DAAC**